

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Air Pollution Monitoring and Prediction

Sheikh Saeed Ahmad, Rabail Urooj and
Muhammad Nawaz

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/59678>

1. Introduction

One of the most important emerging environmental issues in Asian cities is air pollution. Air pollution is an atmospheric condition in which the concentration and duration of certain substances present in the air produce injurious and destructive effects on both man and the surrounding environment [1]. The most common pollutants in air are sulfur oxide, nitrogen dioxide, carbon monoxide and dioxide, and particulate matter.

Geographical Information Systems (GISs) are computer-based applications used for mapping and analyzing the earth and related spatially distributed phenomena. GIS applications integrate unique visualizations with common databases, which make it possible to capture, model, manipulate, retrieve, analyze, and present the geographically referenced data. Compared to other information systems, GIS systems have advantages, including the high power of analyzing spatial data and handling large spatial databases.

GIS applications can be used in air quality management and for controlling pollution, for handling and managing large amount of data. GIS systems manage spatial and statistical data, which facilitates depiction of the association between the frequency of human activities leading to bad environmental health and poor air quality. GIS modeling and statistical analysis also enables to examine and predict the impact of climatic variables on air pollution. In this way, GIS systems help in monitoring air pollution and emissions of pollutants from different sources.

Air pollution mapping is a helpful method for determining the concentration of pollutants. As the result of air pollution mapping, overviews of pollution in cities can be created and their sources of pollution emission can be identified, which help in controlling emissions. Different studies have been executed on air pollution in conjunction with GIS [2-11]. Consequently, GIS

applications in air monitoring are necessary to determine air quality to reduce pollution to such a level at which harmful impacts on human health and the environment is reduced.

With the help of GIS applications, an output report of pollutants in Air Quality Management Systems (AQMSs) can be achieved in the form of three-dimensional (spatial) records. In AQMS emission time, concentration and place of air pollutants are regulated in order to achieve the predefined air quality standards of ambient air. It encompasses the estimation of the pollutants' emission schedule in a way to determine the consequences to air quality and the design of alternative programs for emission control in order to meet air quality standards, which are subject to some limitations, for example, technological viability and lowest charges. For environmental modeling with GIS applications, AQMSs are considered to locate monitoring stations, for development of geospatial model for air quality, and for spatial decision-support systems. However, the most significant step in an AQMS is data mining. The data mining method is a skill, which is used to analyze the data, uncover hidden patterns, and find interesting information from large amounts of data or huge databases. The most commonly used technique in data mining is artificial neural networks [12].

The human brain consists of a large number of neurons connected to each other by synapses to make networks, and these networks of neurons are called neural networks, or natural neural networks. Similarly, the artificial neural network (ANN) is basically a mathematical model of a natural neural network. The ANN uses a mathematical or computational model based on connectionist approach for solving the given problem. The concept of ANN is derived from biological neural network systems. The key applications of neural networks are control systems, classification systems, and prediction and vision systems.

Three basic components are important in order to make functional model, like: synapses of neuron; an added that sum all input in form of weights; and activation function. In Figure 1, synapses are shown by weights. Basically, a strong connection between input and neuron is noted by synapses or value of weight. Negative values reflect inhibitory connections, whereas excitatory connections are shown by positive values. Activation functions regulate the output of neurons within an acceptable range from -1 to 1.

1.1. Sources

Air pollution takes place due to natural and anthropogenic activities. But air pollution as the result of man-made activities like fossil fuel combustion, construction, mining, agriculture, and warfare are the most significant and cause problems in the atmosphere [13].

Basically, two types of pollution sources have been categorized, i.e., Stationary and Mobile. The stationary source is a type of source that is fixed or is a preset pollutant emitter, for example, fossil fuel burning power plants and refineries. The mobile source is a nonstationary type of pollutant emitter, for example, vehicles. The most emerging and leading cause of air pollution is the motor vehicle [14]. Pollutants that are emitted directly from the source into the air are known as primary pollutants, for example, carbon dioxide, carbon monoxide, sulfur dioxide, etc. When these primary pollutants react in atmosphere with each other to form another type of pollutants, they are called secondary pollutants, which are not directly emitted

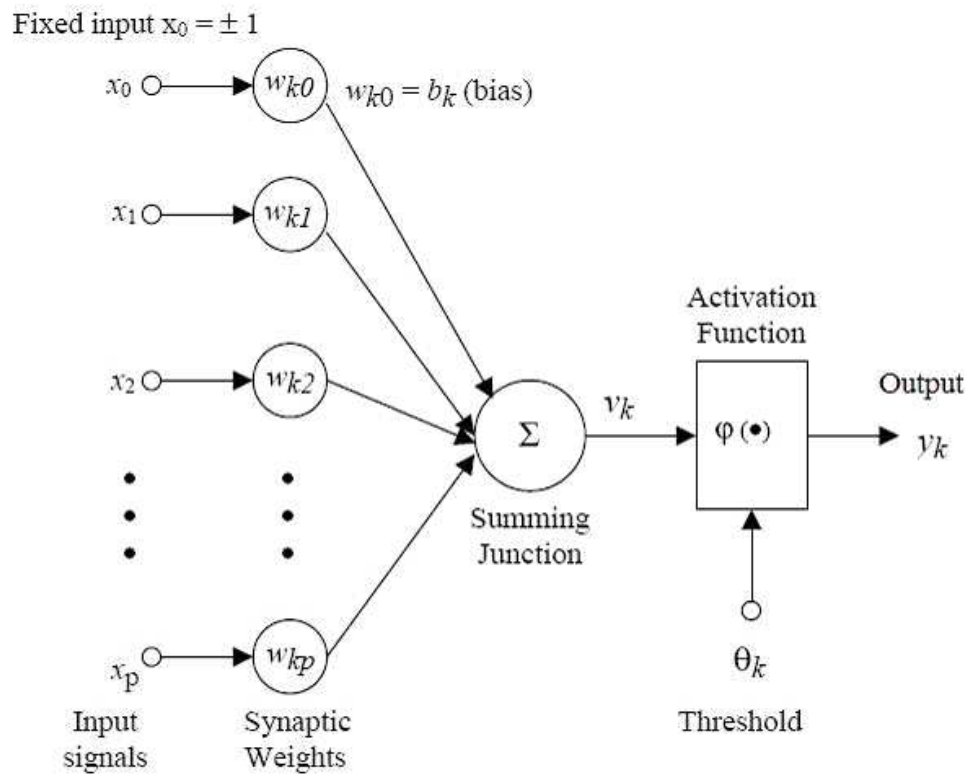


Figure 1. Model of a neuron

but formed as a result of primary pollutants' reaction in the atmosphere. For example, ozone forms when nitrogen oxides react with hydrocarbons in the presence of sunlight, and the resulting nitrogen dioxide reacts further with oxygen and forms ozone as pollutant.

1.2. Health effects

Air pollution and its effects in rural and urban areas are directly related to the ongoing activities. For example, in cities, pollution is related to the products of combustion in industries and vehicles. Many large cities all over the world exhibit excessive levels of air pollutants. Among all dangerous pollutants, nitrogen dioxide (NO_2) is important due to its capacity of causing dangerous effects on humans and the environment, which results in photochemical oxidation and acid rain.

The effects of air pollution cannot be ignored even within homes. Many air pollutants can cause cancer and other diseases among inhabitants. In 1985, it was reported that indoor toxic chemicals are three times more potent in causing cancer than outdoor air pollutants [15]. In America, health issues caused by buildings are called "sick building syndrome"[16].

1.3. Case study

In Pakistan, air pollution is emerging as a serious problem in its mega cities, which needs to be monitored and addressed at the root level in order to reduce the lethal impacts of pollutants

on man and environmental health. The present study of Pakistan focuses on the most important twin cities of Pakistan, which are Rawalpindi and Islamabad. Both cities are commonly viewed as one unit and are 15 km apart. The study area with 135 sampling locations is shown in Figure 2. The climatic condition of Rawalpindi and Islamabad is sub-humid to tropical, with hot and long summers (May to August) accompanied by a monsoon season (July to August) followed by short and mild winters (October to March). The average low temperature is 12.05 °C in January and average high temperature is 31.13 °C in July.

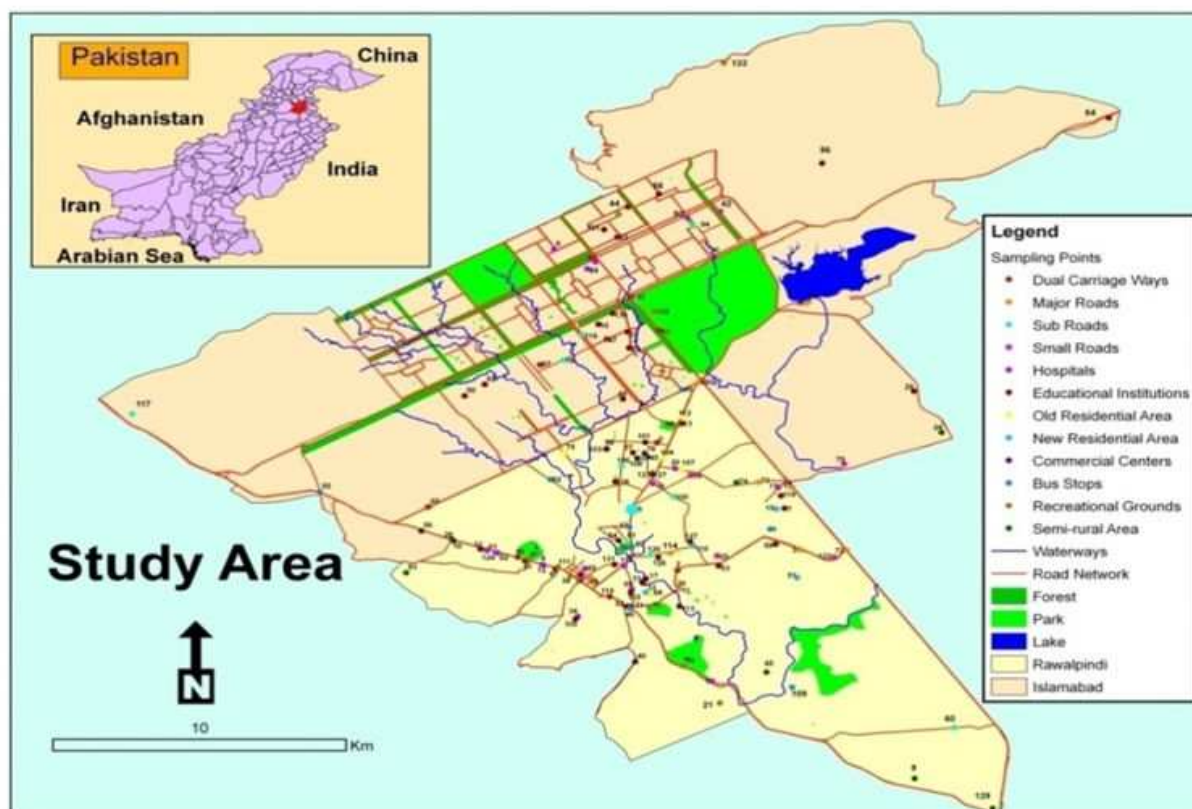


Figure 2. Base map

For the monitoring campaign, the maximum area (135 sampling sites) was covered in order to represent different traffic intensity and congestion levels in the urban area of Rawalpindi and Islamabad, for sampling. These sites included dual carriageways, major, linking, and small roads, healthcare centers, educational institutes, commercial areas, old residential areas, modern residential areas, recreational spots and semi-rural areas.

Research was carried out in order to monitor the NO_2 concentration in the ambient air of Rawalpindi city. Passive samplers were used within the city from January to December in 2008. The average concentration found was 27.46 ± 0.32 ppb. The highest concentration was recorded near the main roads and in the vicinity of schools and colleges due to the large number of transport vehicles, which exceeded the set limit concentration value given by the World Health Organization.

2. Experimental design

2.1. Passive Sampling of NO₂

The most frequent method in monitoring studies for passive sampling of NO₂ is using diffusion tubes described by Atkins [17]. This method for NO₂ measurement is reliable, easy to handle, and it is an inexpensive method for screening air quality. Moreover, passive samplers are preferably appropriate for extensive spatial measurement of NO₂, and they have been reported in many studies of NO₂ monitoring of air in many countries like the United Kingdom, USA, France, Turkey, Argentina, and China [18].

Basically, passive samplers are designed on the principle of air diffusion having an efficient absorber at one end of the tube, and the flow rate (sampling rate) at constant temperature can be measured by using Flick's Law [19]. For that, the length and diameter of diffusion tubes are known, whereas sampling by using diffusion tubes is independent of air pressure.

2.2. Neural network design

From different sampling sites covering the whole study area, data was collected for neural network analysis. Collected data was fed to the neural network that has area_id, season_id, temperature, humidity, rainfall, and the respective concentrations as columns. For the neural network, the marked value was set to predict concentrations and rests were used as input to the neural network.

Neural network has two phases: training and testing. In the first phase (training), the network is trained by providing the complete information about the characteristics of data and observable outcomes to perform a particular task.

A neural network can develop a model that learns the relationship between input data and the desired outcome in the training phase. In the testing phase, testing data are provided as input. The performance of the testing phase depends upon the training phase (it depends on the number of samples that are provided during the training phase and also on the number of times that the network is accurately trained. However, it is impossible that the output is 100% precise for any network input. MS Access was used as the database engine because it is easy to use for all.

For testing the neural network, the cross validation method is used by using holdout method in which data was divided into testing and training data. The database consisted of two tables: training_data and testing_data. The function of training_data is to train the ANN by adjusting weights in order to maximize the predictive ability of ANN and minimize error during forecasting. Testing data was used to test the prediction accuracy of ANN on new data. The structure of training data and testing data is given in Table 1.

In Table 1, the first key "id" is primary key, which contains the number that indicates row number and the second key "loc_id" contains the number that indicates location from where data is gathered, loc_name indicates the name of location and the next six fields indicate

position of location with respect to north and east. The next two indicate temperature and humidity levels.

The 13th and 14th fields indicate concentration of NO₂ and level of concentration value. The last field of dataset contains week number, which indicates the number of weeks in which data is gathered from particular location. The attribute for testing data are the same in the testing data structure.

Field Name	Data type	Primary key	Field size
Id	Number	Yes	Long Integer
loc_id	Number		Long Integer
loc_name	Text		50
map_id	Number		Long Integer
north_d	Number		Long Integer
north_m	Number		Long Integer
north_s	Number		Long Integer
east_d	Number		Long Integer
east_m	Number		Long Integer
east-s	Number		Long Integer
Temp	Text		50
Humidity	Text		50
Concentration	Number		Long Integer
con_level	Number		Long Integer
Week	Number		Long Integer

Table 1. Structure of training data

For designing a network, we need to specify the architecture of a neural network by designing a number of hidden layers and units in each layer along properties of network that describe error function and network activation.

For optimal generalization of collected data, two types of architectures: the rtNEAT (real-time neuro evolution of augmented topologies) architecture with evolution algorithm and the feed forward architecture with back propagation algorithm of ANN are used in order to ensure high accuracy of ANN prediction about impacts of NO₂ concentration achieved in future. This rtNEAT architecture is used to train neural network with evolutionary algorithm, which has three steps, i.e., selection, mutation, and reinsertion. But before the training of neural network, the topology has to be created in the design of the neural network. A neural network is a connection of neurons, which contains three types of nodes: input, output, and hidden node. All nodes are randomly created during its execution.

Table 2 describes the properties of network, which contains an error function and network activation parameters. These properties are functional to all tested networks by the architecture search method and manually selected network.

Parameter	Value
Input activation FX	Logistic
Output name	Concentration
Output error FX	Sum-of-squares
Output activation FX	Logistic

Table 2. Network properties

The logistics function has a sigmoid curve and sum of squares. The sum of squares is the most frequent function error, which is used for the classification problem. The error is the sum of the square differences between the real input value and neural network target value.

2.3. Architecture search

A heuristic search is used to search the dataset for the best networks. Heuristic methods are used to speed up the process of finding a satisfactory solution. The architecture search for the designed neural network NO₂ is given in Table 3.

ID	Architecture	# of Weights	Fitness	Train Error	Validation Error	Test Error	AIC	Correlation	R-Squared
1	[8-1-1]	11	0.079965	11.371084	12.151164	12.505404	-4220.122343	0.652911	0.424949
2	[8-20-1]	201	0.080369	11.295746	12.093373	12.442678	-3846.484024	0.653839	0.427147
3	[8-12-1]	121	0.080841	11.017718	12.044772	12.37002	-4030.333807	0.668927	0.446734
4	[8-7-1]	71	0.080593	11.182193	12.108417	12.407984	-4116.153129	0.662147	0.438331
5	[8-16-1]	161	0.081507	10.941978	11.986108	12.26894	-3956.935279	0.670637	0.448777
6	[8-18-1]	181	0.080474	10.917611	12.026946	12.4264	-3919.068903	0.676823	0.457557
7	[8-14-1]	141	0.080839	11.105445	12.044827	12.370266	-3982.744071	0.666178	0.44366

Table 3. Heuristic architecture search for NO₂

2.4. Training of neural network

The next step is to train the neural network for the NO₂ dataset by using the propagation algorithm. Weight change is calculated by the quick propagation algorithm by utilizing the quadratic function $f(x) = x^2$. In neural networks, several layers contain neurons in each layer that are connected with each other like neurons in the input layer connected to one or more

neurons of the hidden layer, which are further connected to the output layer's neuron. With each presentation in neural network, error is computed as the difference between network output and observable output. The combination of randomly assigned weight (giving low error) replaces weights that are at the first location. This is called training to adjust the connection weights to enable the network to produce the expected output. Two different weights having two different error values are two points of a secant. Relating this secant to a quadratic function, it is possible to calculate its minimum $f(x) = 0$. The x-coordinate of the minimum point is the new weight value.

$$S(t) = \frac{\partial E}{\partial w_i(t)}$$

$$\Delta w_i(t) = \alpha \cdot \frac{\partial E}{\partial w_i(t)} \text{ (Normal back propagation)}$$

$$\frac{\Delta w_i(t)}{\alpha} = \frac{\partial E}{\partial w_i(t)}$$

$$S(t) = \frac{\partial E}{\partial w_i(t)} = \frac{\Delta w_i(t)}{\alpha}$$

$$\Delta w_i(t) = \frac{S(t)}{S(t-1) - S(t)} \cdot \Delta w_i(t-1) \text{ (Quick propagation)}$$

Here w =weight, i =neuron, E =error function, t =time (training step), α = learning rate, and μ = maximal weight change factor

The quick propagation coefficient was set to 1.75, learning rate was 0.1, and iterations were 500. The training graph for dataset errors for NO₂ is shown in Figure 3.

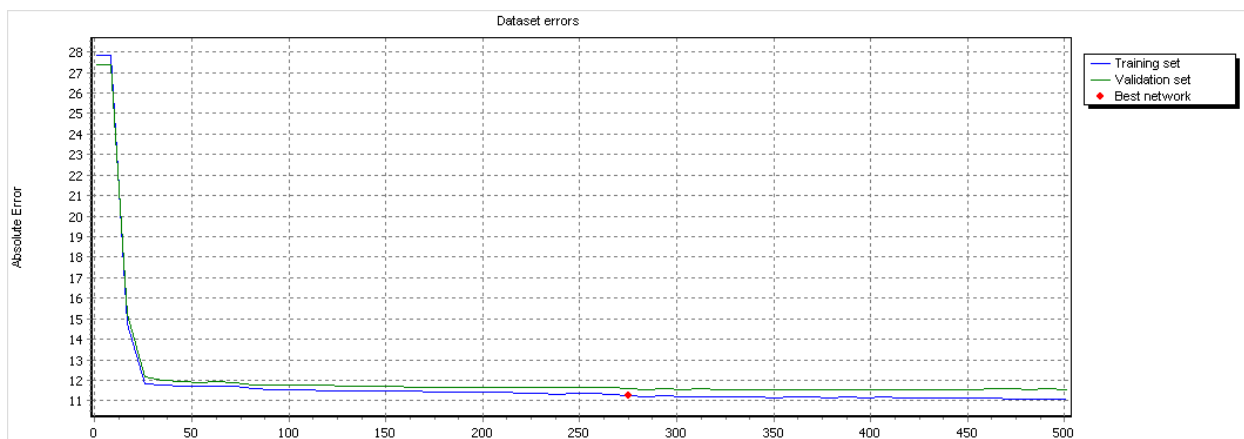


Figure 3. Dataset errors for the NO₂ dataset

The training graph of correlation for NO₂ is shown in Figure 4.

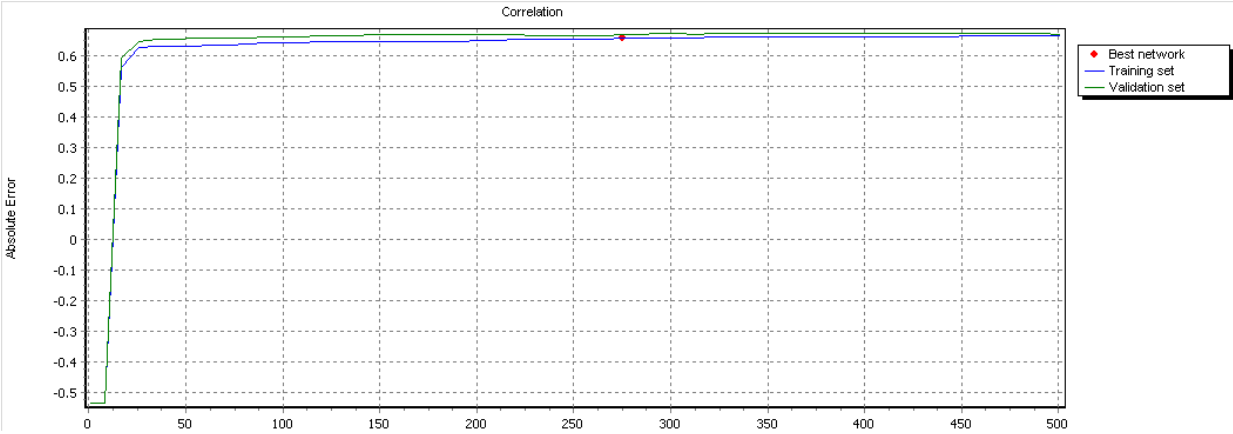


Figure 4. Graph of correlation for NO₂

The graph of error improvement – network errors for NO₂ is shown in Figure 5.

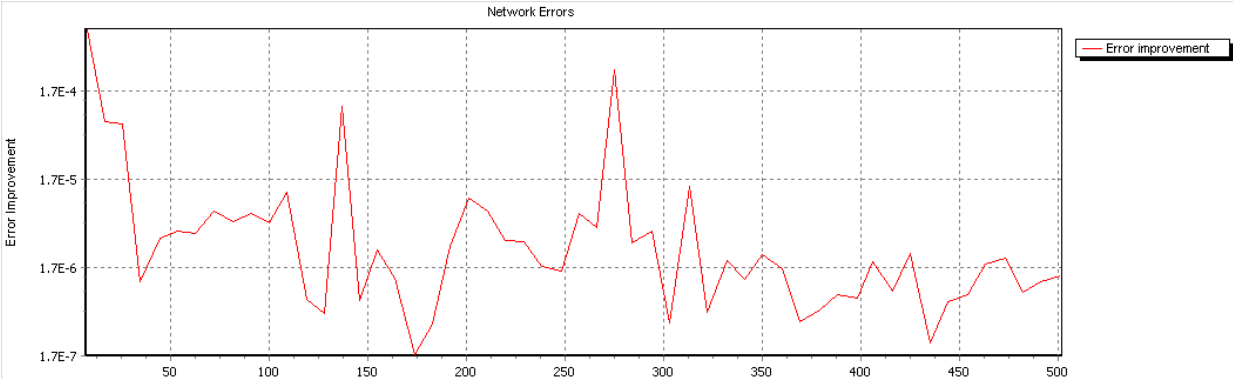


Figure 5. Network errors for NO₂

The error distribution of network statistics obtained after training of neural network is shown in (Figure 6).

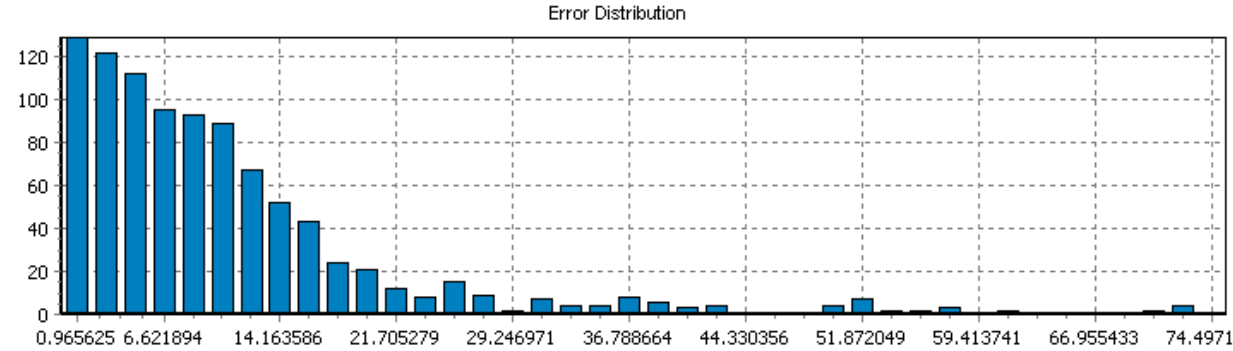


Figure 6. Error distribution for NO₂

3. Data analysis

In order to determine the seasonal variation and statistical significance, results are presented in tabular format. Tables 4 a and 4 b show the average concentration level of NO₂, season-wise, along standard deviation (SD) values measured at different sampling sites of study.

Table 4 a shows average values of NO₂ concentration in different seasons of 12 major sampling categories in urban Rawalpindi and Islamabad from November 2009 to July 2010.

Table 4 b shows the seasonal average concentration of NO₂ of 12 major sampling categories in urban Rawalpindi and Islamabad from September 2010 to March 2011.

Table 5 presents NO₂ concentration for each selected category, as described in study area profile, to understand the general trends of NO₂ concentration levels among different categories during the course of experimental period.

Sampling Categories	Mild Winter (Nov)	Winter (Dec to Jan)	Early Spring (Feb)	Spring (Mar)	Mild Summer (April)	Summer (Pre-Monsoon) (May to June)	Monsoon (July to August)
	NO ₂ Conc. (weekly basis) (ppb)	(ppb)	(ppb)	(ppb)	(ppb)	(ppb)	(ppb)
Rawalpindi	Dual Carriage Ways (5)	87±19.78	98±26.87	63±12.29	53±6.49	44±10.64	22±4.22
	Major Roads (10)	60±12.19	68±9.56	52±13.52	45±10.23	36±8.97	26±5.88
	Sub-roads (6)	74±20.50	86±24.47	60±16.49	50±11.05	38±12.65	33±13.01
	Small Roads (3)	55±9.78	63±4.89	47±5.57	40±8.24	31±3.40	25±4.68
	Public Hospital (5)	48±18.71	63±18.40	37±0.74	29±2.29	22±2.24	18±0.79
	Private Hospitals (8)	61±14.47	75±14.19	38±1.16	32±2.03	25±2.29	20±5.57
	Public EI (11)	85±30.58	95±32.94	75±23.75	63±17.94	47±17.37	31±10.14
	Private EI (17)	55±9.71	66±9.54	45±4.56	43±9.65	38±10.89	26±4.54
	Old Residential Areas (5)	83±15.24	95±16.09	55±13.32	51±6.66	37±6.44	26±2.54
	Modern Residential Areas (5)	65±20.07	73±14.89	69±24.49	59±12.55	36±7.13	28±5.08
	Commercial Area (2)	75±0.83	82±17	61±6.69	51±7.11	36±4.29	21±6.20
	Bus Stops (9)	74±20.26	83±31.47	69±33.78	58±17	39±17.32	28±8.41
	Recreational Spots (9)	75±38.40	87±40.76	62±36.39	56±21.88	43±19.97	31±11.12

Sampling Categories		Mild Winter (Nov)	Winter (Dec to Jan)	Early Spring (Feb)	Spring (Mar)	Mild Summer (April)	Summer (Pre-Monsoon) (May to June)	Monsoon (July to August)
NO ₂ Conc. (weekly basis)		(ppb)	(ppb)	(ppb)	(ppb)	(ppb)	(ppb)	(ppb)
Islamabad	Dual Carriage Ways (3)	84±28.73	95±33.64	66±23.78	57±12.31	45±16.69	24±5.98	19±4.16
	Major Roads (3)	50±3.72	60±2.04	40±0.81	32 ±2	26±4.42	21±2.97	15±2.16
	Sub-roads (4)	54±6.06	67±6.39	49±6.49	43±7.24	38±12.79	25±1.38	18±1.26
	Small Roads (3)	59±12.65	64±6.33	51±9.60	44±8.93	35±4.53	26±3.66	20±3.08
	Public Hospitals (3)	44±0.58	57±0.29	39±0.29	32±0.58	23±1.47	19±0.51	15±1.71
	Private Hospitals (1)	42	56	38	30	24	19	14
	Public EI (5)	53±13.34	64±9.32	46±9.30	39±10.76	34±14.19	25±6.01	18±1.28
	Private EI (6)	58±11.23	63±7.18	49±7.72	39±9.93	31±5.85	24±2.66	17±1.77
	Commercial Area (1)	61	68	57	50	35	25	16
	Bus Stops (12)	72±14.25	78±16.23	65±7.51	55±5.23	34±6.22	25±3.21	19±2.56
	Recreational Spots (2)	62±5.97	69±4.58	57±2.45	48±1.59	38±2.48	25±3.15	17±1.56
	Semi-Rural Areas (7)	46±8.98	59±5.64	42±6.41	33±7.87	31±5.29	24±3.22	18±3.19
(a)								
Rawalpindi	Dual Carriage Ways (5)	30±4.00	51±7.52	88±22.28	100±26.42	63±18.52	50±20.32	
	Major Roads (10)	27±6.57	49±9.72	61±10.33	68±9.34	48±3.76	37±3.79	
	Sub-roads (6)	32±7.31	53±12.97	74±23.42	87±26.60	58±13.69	39±6.36	
	Small Roads (3)	28±3.65	37±2.53	53±6.30	62±2.58	54±12.39	43±11.13	
	Public Hospital (5)	20±0.98	32±5.46	48±18.01	64±18.03	40±6.13	29±3.94	
	Private Hospitals (8)	23±3.90	38±7.19	60±15.37	73±14.03	40±4.05	31±1.95	
	Public EI (11)	44±16.98	81±36.87	86±31.78	96±34.20	73±21.26	63±18.23	
	Private EI (17)	31±4.85	42±6.10	55±8.91	66±9.82	45±7.08	35±5.90	

Sampling Categories		Mild Winter (Nov)	Winter (Dec to Jan)	Early Spring (Feb)	Spring (Mar)	Mild Summer (April)	Summer (Pre-Monsoon) (May to June)	Monsoon (July to August)
NO ₂ Conc. (weekly basis)		(ppb)	(ppb)	(ppb)	(ppb)	(ppb)	(ppb)	(ppb)
Islamabad	Dual Carriage Ways (3)	31±5.72	50±11.40	82±21.11	99±32.70	67±19.78	49±12.84	
	Major Roads (3)	22±0.80	37±1.93	53±4.33	65±0.30	44±2.30	33±2.00	
	Sub-Roads (4)	26±2.48	39±4.60	54±5.74	65±4.08	46±3.02	35±9.17	
	Small Roads (3)	30±5.94	41±4.12	54±4.24	63±6.67	47±2.60	38±2.79	
	Public Hospitals (3)	22±2.14	34±2.66	45±0.80	60±1.41	45±0.22	34±0.82	
	Private Hospitals (1)	22	31	40	55	38	30	
	Public EI (5)	31±7.41	40±3.60	52±7.94	64±8.39	46±9.34	37±9.62	
	Private EI (6)	29±11.58	41±8.65	54±10.14	63±7.03	47±7.93	36±9.17	
Twin Cities	Old Residential Areas (5)	27±2.97	61±14.74	84±14.18	95±16.51	58±12.41	48±10.06	
	Modern Residential Areas (5)	32±7.86	49±11.70	66±20.07	75±16.16	60±19.16	48±16.53	
	Commercial Area (3)	32±1.23	46±6.09	63±1.00	71±3.57	56±7.02	48±8.41	
	Bus Stops (11)	32±9.11	53±20.30	76±20.07	87±32.40	69±31.34	54±19.54	
	Recreational Spots (10)	37±18.55	52±25.23	71±37.63	84±39.83	57±29.71	46±24.78	
	Semi-Rural Areas (7)	31±9.47	41±7.44	53±6.51	62±6.21	44±7.50	36±6.99	
(b)								

Table 4. (a): Seasonal mean values of NO₂ from November 2009 to July 2010 (b): Seasonal mean values of NO₂ from September 2010 to March 2011

In Table 5 most of the sampling sites of study area showed nearly similar average concentration from month of November 2009 to March 2011. Maximum concentration of NO₂ shown on dual carriage ways.

The possible cause of such elevated levels of NO₂ concentration is extensive increase in number of vehicles, increase in population, busy roads, fuel inefficient vehicles, driving ways, and traffic jams. Gilbert reported that NO₂ is considerably related to both the distance from the nearest highway and the traffic count on the nearest highway [20].

The rest of the categories showed nearly the same average concentration. Major roads and sub-roads showed average NO₂ concentration levels of 53.56 ppb and 51.78 ppb, respectively. Sub-roads, bus stops, recreational spots, and educational institutions showed similar concentration levels of approx. 51 ppb.

Sampling Categories	No. of Sites	Average NO ₂ Conc. (ppb)
Dual Carriage Ways	8	55.23
Major Roads	13	53.56
Sub-roads	10	51.78
Bus Stops	11	51.62
Educational Institutions	39	51.26
Recreational Spots	10	51.18
Old Residential Area	5	48.97
Small Roads	6	48.23
Commercial Area	3	47.59
Hospitals	17	47.44
Modern Residential Area	5	46.25
Semi-Rural Area	7	37.65

Table 5. Average NO₂ concentration levels in twin cities from November 2009 to March 2011

Educational institutions and recreational spots, being present close to the dual carriage ways, also experience elevated concentration levels. Old residential areas (48.97 ppb) showed slightly higher NO₂ concentration levels as compared to modern residential areas (47.59 ppb).

Narrow road, enclosing architecture, and congestion among the old residential areas result in traffic emission being trapped and buildup leading to higher NO₂ concentration levels, whereas in modern residential areas increased vehicular number is the major cause of elevated NO₂ levels. The minimum NO₂ concentration levels were indicated in semi-rural areas, that is 37.65 ppb. A study in Vilnius commented the same phenomena; NO₂ average rates depend upon traffic and are highest in cross roads and lowest at the background suburban areas [21].

For annual average concentration level of nitrogen dioxide, a spatial interpolation map has been developed by using inverse distance weighted (IDW). IDW in Figure 7 is clearly depicted as the areas of higher and lower concentration level of NO₂ in Rawalpindi and Islamabad.

Higher concentration levels are represented by darker shades while the lower concentration levels are shown with lighter shades. The maximum NO₂ values were found at the center of the city, where they reached the concentration of 83–110 ppb. Values were low on the outskirts of the city, with the lowest concentration in north (31–44 ppb).

A study in Vilnius commented the same phenomena; NO₂ average rates depend upon traffic and are highest in cross roads and lowest at the background suburban areas. Dual carriage ways, sub roads, major roads, commercial areas, old residential areas, and areas where schools and colleges are existing have higher concentration levels of NO₂. Intense traffic flow and congestion were the major reasons for these elevated levels of nitrogen dioxide concentration in those areas as vehicular emission is the predominant source of NO₂.

Vehicle growth rate in twin cities is extensively high. Load of traffic is continuously increasing with growing population rate and demand of motor producing industry. Due to this, traffic congestion is also increasing day by day with growing vehicle population, resulting in highest emission rates per vehicle.

The higher emission rate of NO₂ can also be attributed to the type of fuel and quality of fuel [22]. In Figure 7 Rawalpindi showed more concentration levels than Islamabad due their building patterns.

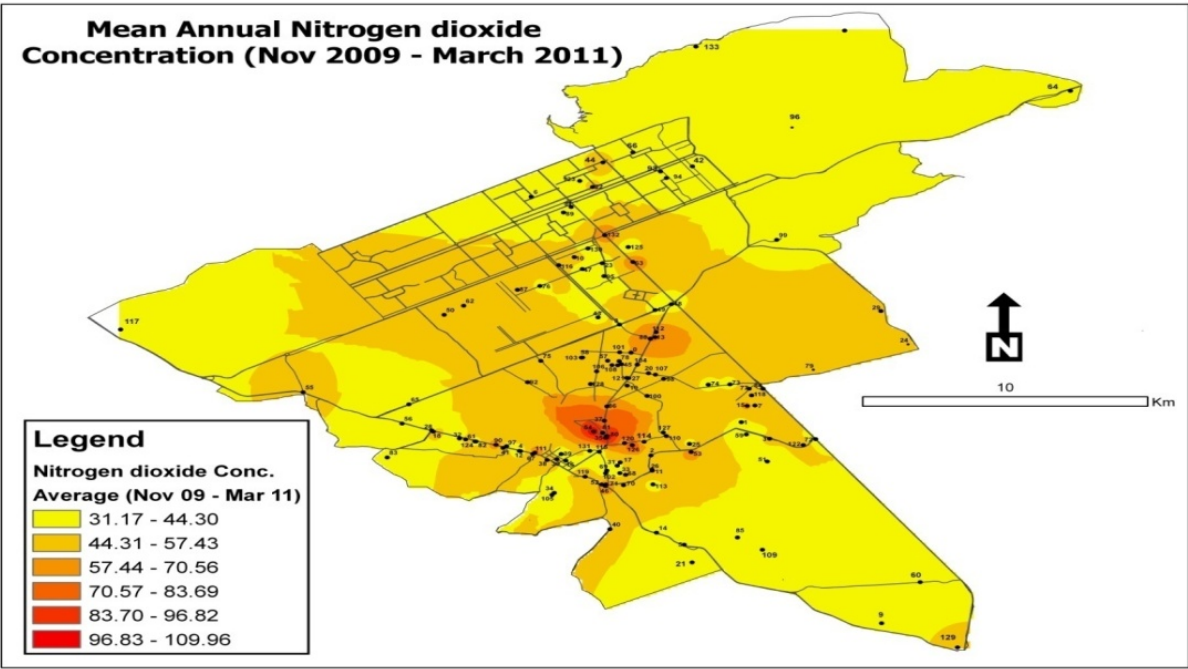


Figure 7. Spatial distribution of NO₂ concentration

3.1. Neural network data analysis

Based on the design of neural network, with the neural architecture and properties discussed, the data space is searched by using heuristic search method with 500 iterations and fitness criteria is set to Inverse Test error. The best top 5 networks explored from the space by the heuristic search are graphically shown (Figure 8).

Heuristic search is a problem-solving method that analytically searches a space of problem states. The best network is obtained when the absolute error gets minimum in the initial iterations so the best network out of the 5 best networks is shown (Figure 9).

Results for all data sets produced after training and testing data. Real vs. target graph represented a line graph of real- and network-predicted target values for record displayed in Table 6. X-axis shows the selected input column values and Y-axis represents network-predicted output values. Table 6 presents the summary of the real vs. output table after training.

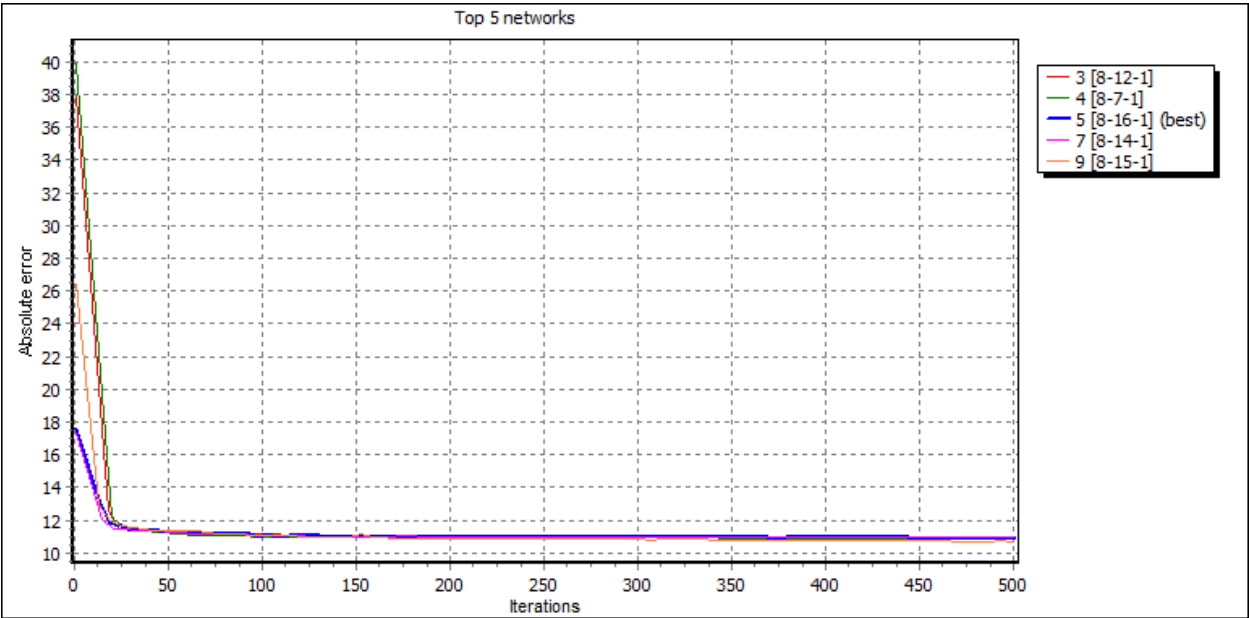


Figure 8. The top five networks explored by heuristic search approach

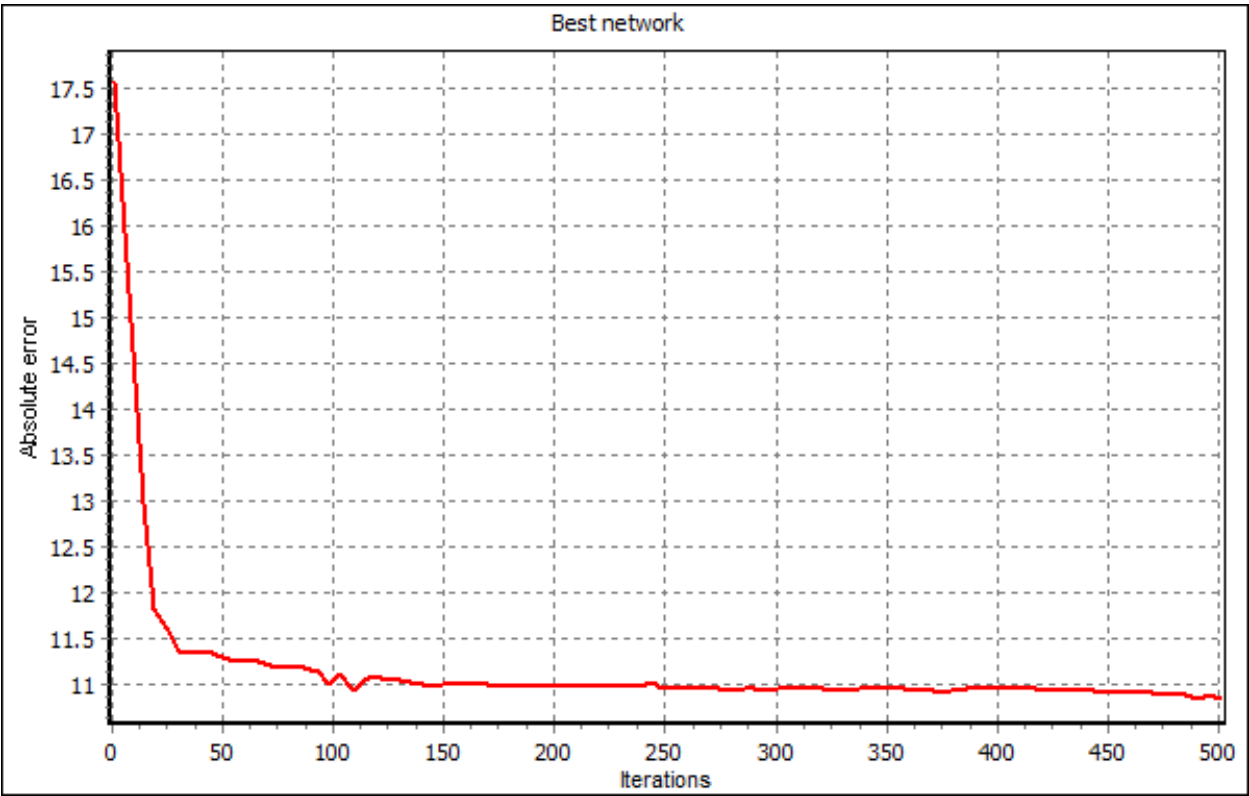


Figure 9. Network explored by heuristic search

	Target	Output	AE	ARE
Mean:	45.237265	45.09091	11.341221	0.250292
StdDev:	20.98552	13.97879	11.112997	0.180871
Min:	11.3	20.16986	0.004569	0.000171
Max:	132.72	63.353673	73.986765	1.096446
Correlation, 0.653989; R-squared, -0.290243				

Table 6. Summary of real vs. target

The visualization for real vs. output with row number on x-axis and target/output (area_id) on y-axis is shown (Figure 10).

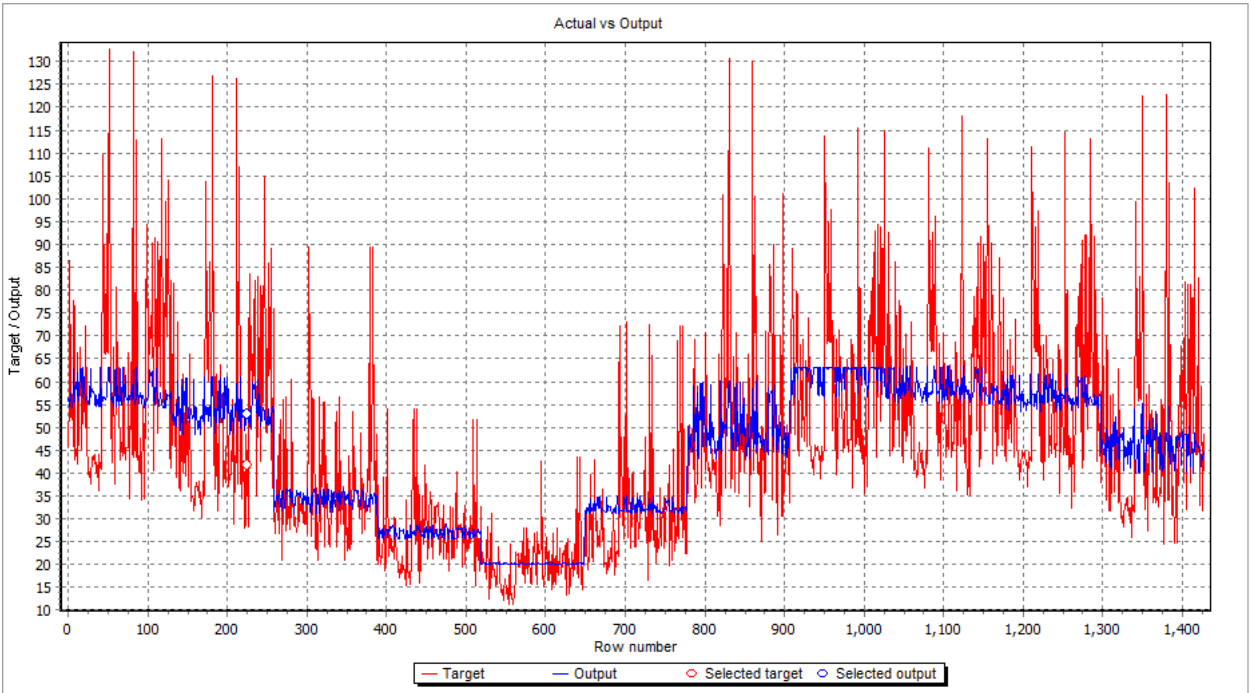


Figure 10. Real vs. network output

Figure 11 shows a scatter plot of the real and forecasted output values. X- axis presents the real values and Y-axis shows predicted network values.

Graph in Figure 12 shows the Network Error Dependence on values, which are numerically input in columns of data sheet. Through graph of Error Dependence, the ranges of the selected input column that can produce network error can be identified.

The last phase after the neural network is trained and tested is to query the network. The concentration is the output value for the neural network. So the input queries are subjected area_id, season_id, temperature, relative humidity, and rainfall (Figure 13).

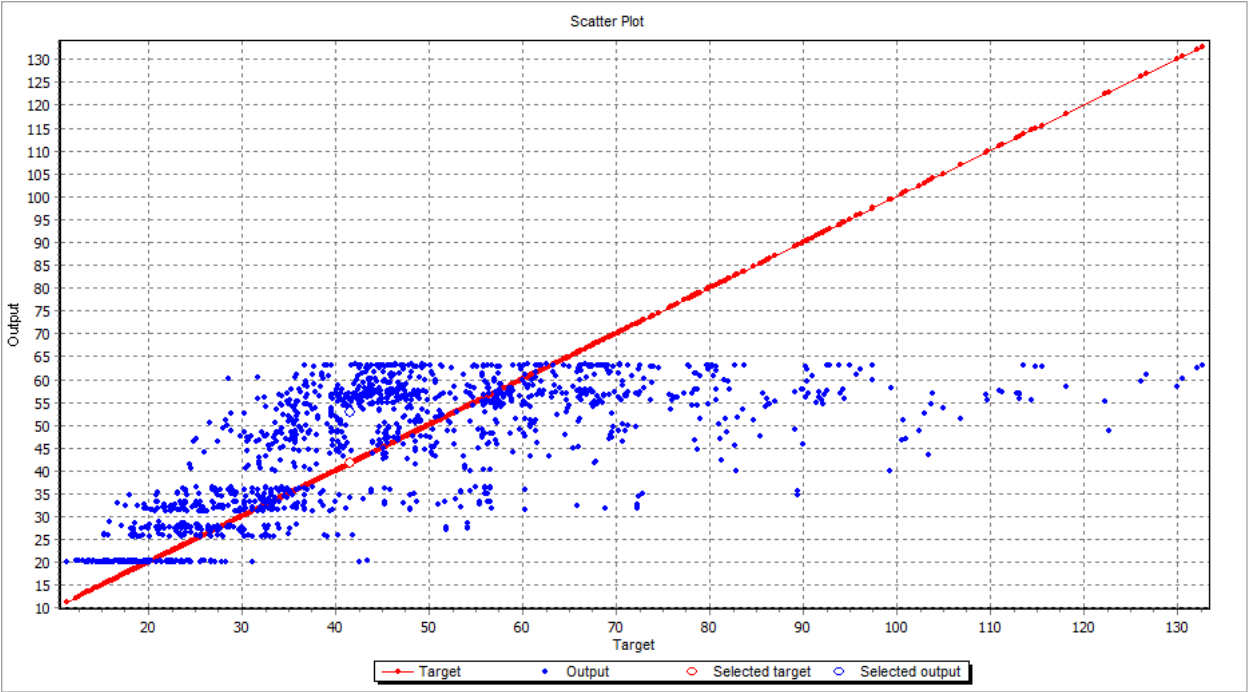


Figure 11. Scattered plot of real and output network values

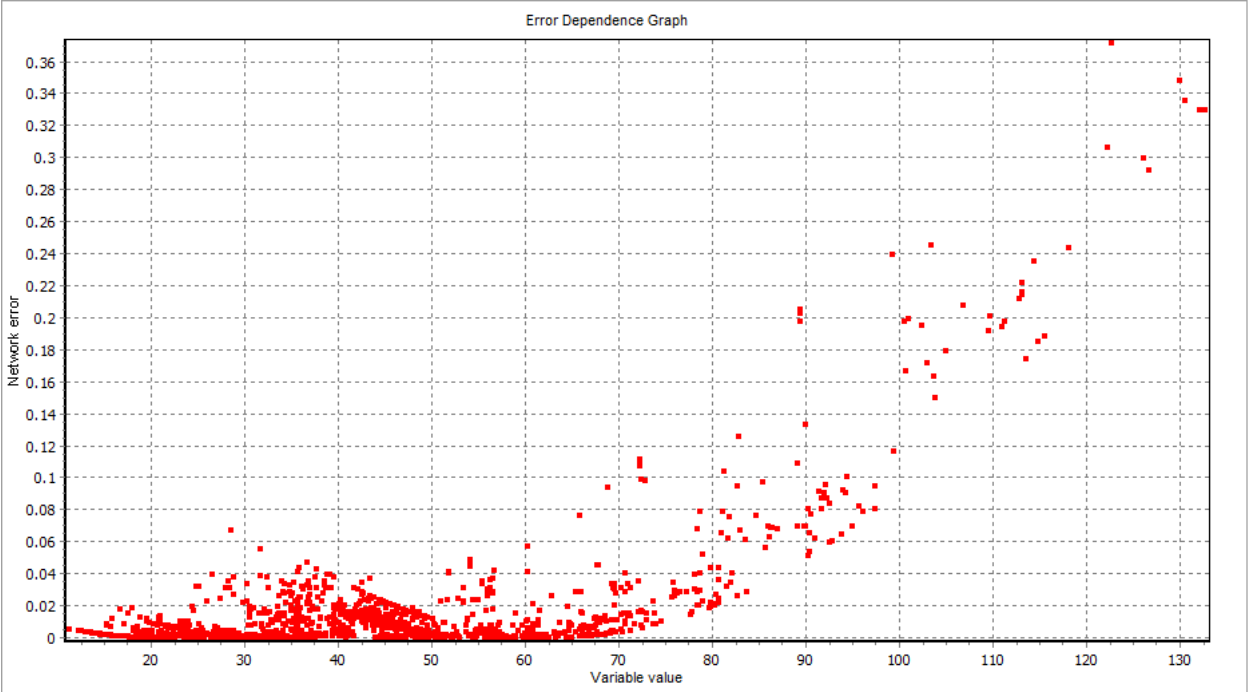


Figure 12. Graph of error dependence

The input Excel sheets are prepared for the GIS mapping. Sheets include area_id, their latitude, longitude, and their concentrations. With the help of interpolation, maps are created for the service.

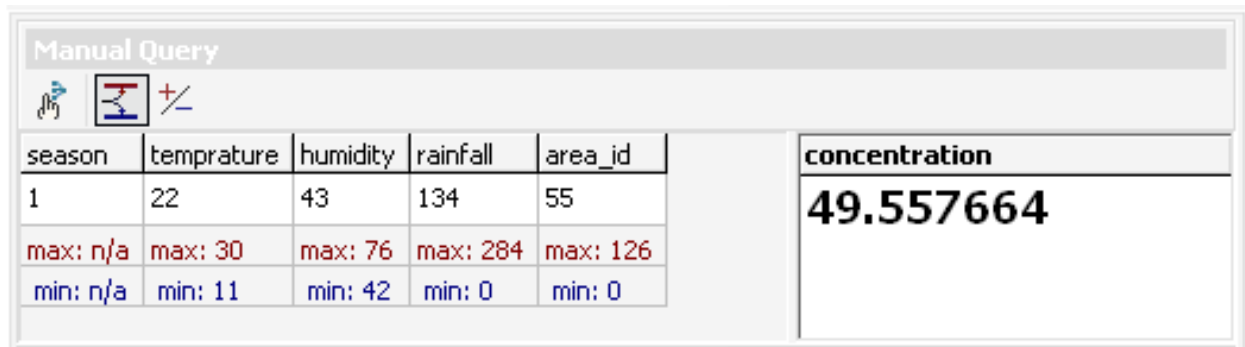


Figure 13. Excel sheet presenting manual query

Temporal variation can be explained through meteorological recorded conditions. However, most of the variations on a local scale are due to the impact of air pollutants.

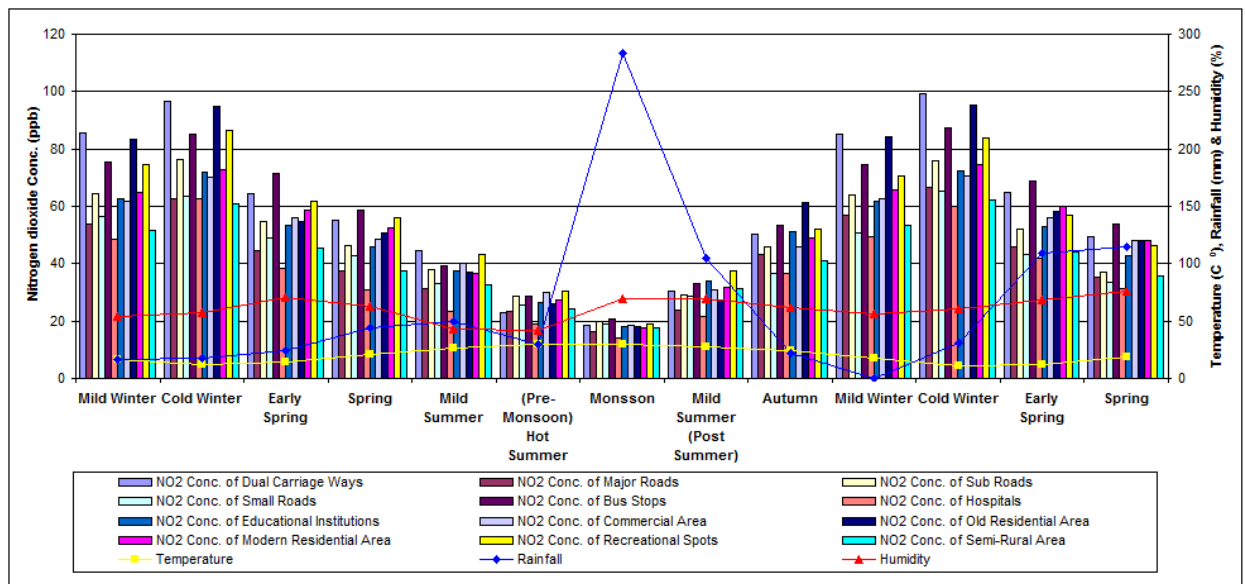


Figure 14. Relationship of rainfall, temperature, and humidity with NO₂ concentration (November 2009–March 2011)

Figure 14 indicates the positive association of NO₂ concentration level with humidity (RH in %) and negative association with the temperature. Figure 15 shows the concentration of NO₂ during summer when recorded temperature, rainfall, humidity are 31°C, 67, and 17mm, respectively.

Figure 16 shows the concentration of NO₂ during the winter season at 11 °C, 68% humidity, and 9mm rainfall.

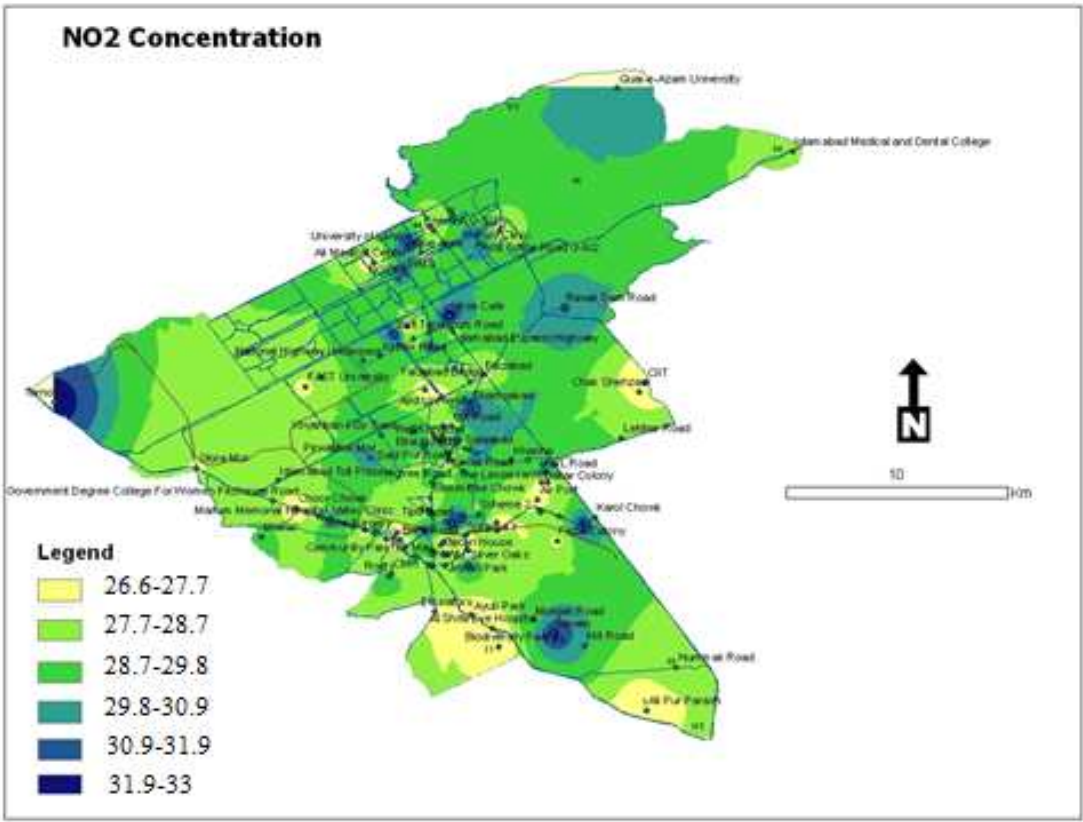


Figure 15. NO₂ concentration in summer

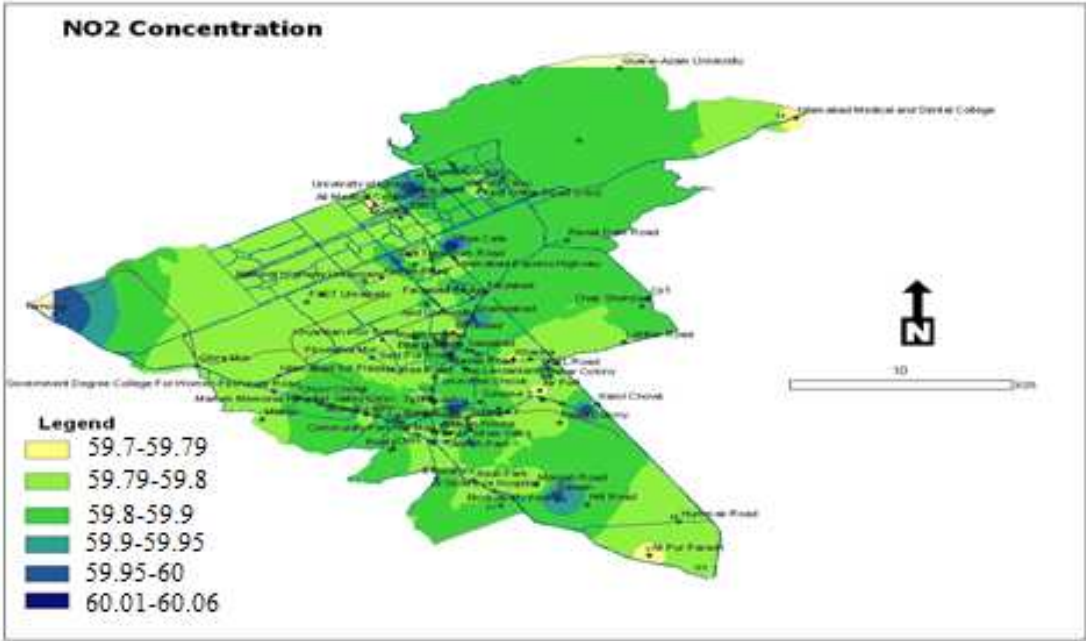


Figure 16. NO₂ concentration in winter

Concentration of NO_2 during the spring season, shown in Figure 17, when recorded temperature is 35°C , humidity is 58%, and rainfall is 60 mm.

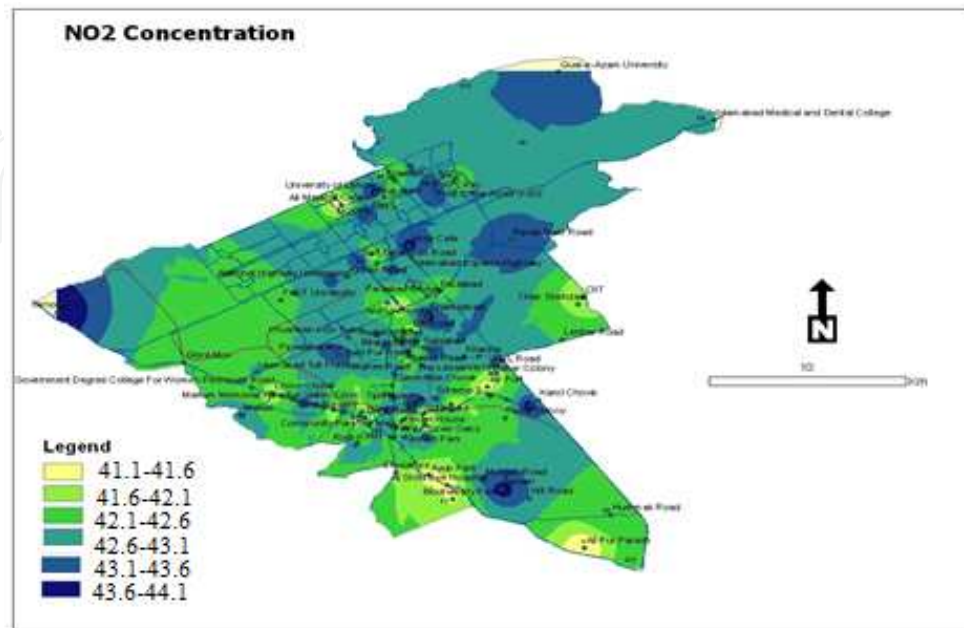


Figure 17. NO₂ concentration in spring

Figure 18 shows predicted concentration of NO₂ in autumn season when recorded temperature, humidity, and rainfall are 29 °C, 69, and 22 mm, respectively.

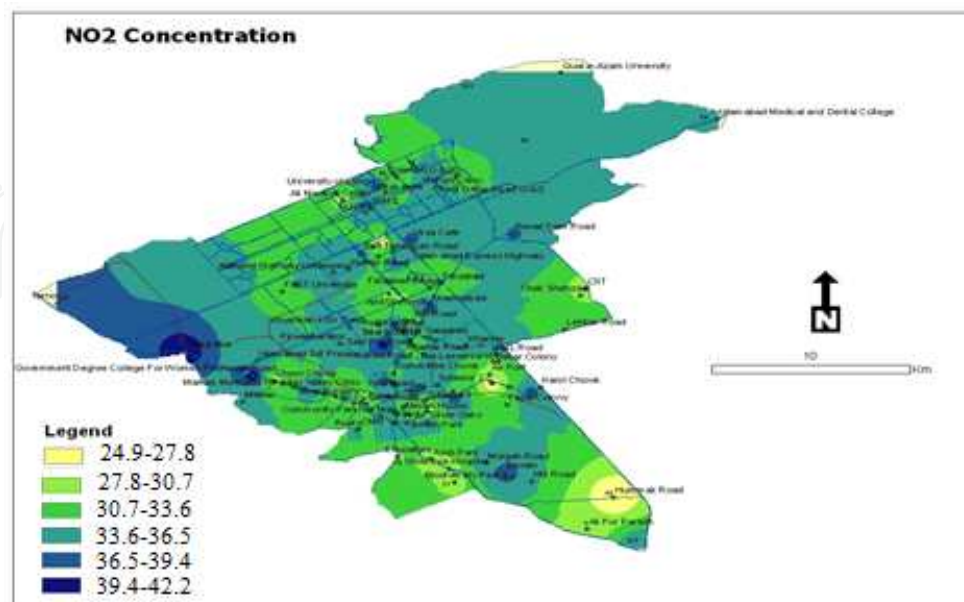


Figure 18. NO₂ concentration in autumn

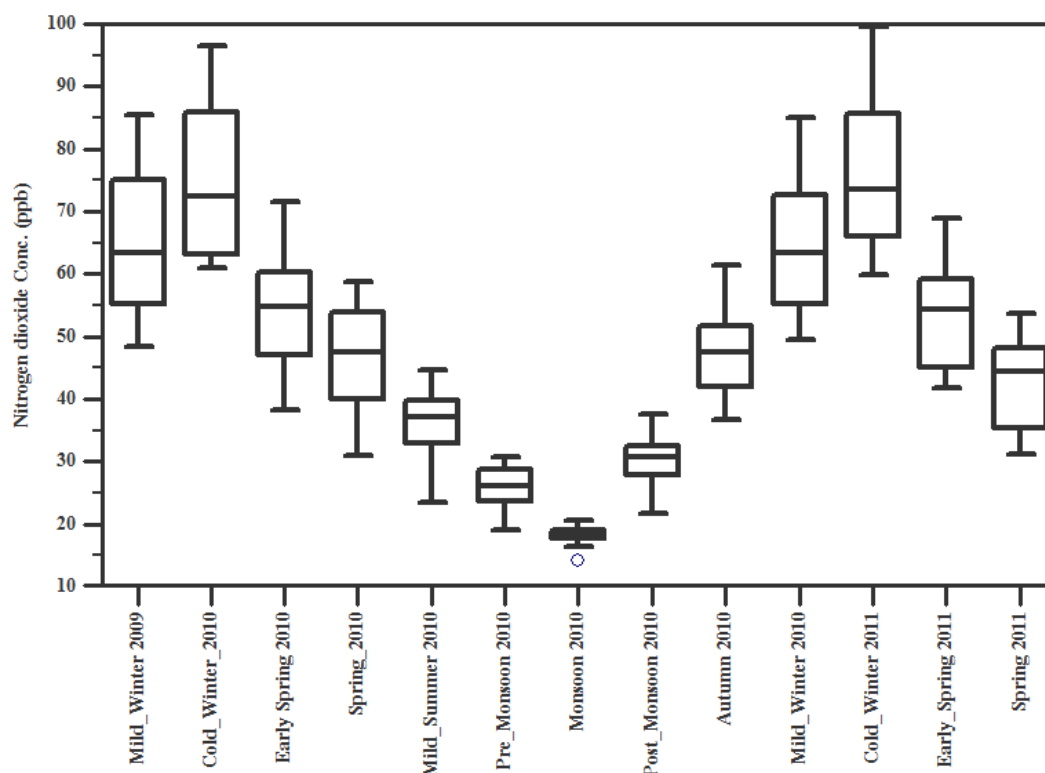


Figure 19. Seasonal variation in NO₂ concentration levels (November 2009 – March 2011)

Figure 19 shows that concentration of NO₂ varies in different seasons. The months from May to August were months in which the minimum value of NO₂ was recorded, and the maximum concentration was measured in the winter season from December to January.

4. Conclusion

NO₂ concentration levels were recorded on hourly and weekly basis in Rawalpindi and Islamabad city by using diffusion tubes. Artificial neural networks were trained to generalize the process of air pollutant spread over three dimensions. Prediction capabilities of ANN were analyzed through generalization by using hold-out evaluation method of classification. Results showed the advantage of using rtNEAT-like architecture of ANN where a neural network can modify its architecture to reduce the error up to the maximum possible limit. Results showed that annual average concentration of NO₂ concentration was 44 ± 6 ppb. However, the highest concentration was recorded in winter season near the dual carriage ways, schools, and colleges because of the higher number of transport vehicles on the road. This endorsed the fact that the reduced photolysis leads to the accumulation of NO₂ during winter due to less solar radiation. This is again attributed by the results of correlation, which reveal the negative correlation of nitrogen dioxide concentration levels with rainfall and temperature and the positive correlation with humidity. Moreover, the results of correlation reveal that the

measured NO₂ concentration levels at different sampling areas exceeded the set limit of concentration value of the World Health Organization and Pak-EPA standard policy. This type of investigative study of artificial neural networks in the area of air pollution modeling shows promising applications for advanced machine learning algorithms in the emerging area of research called eco-informatics.

Author details

Sheikh Saeed Ahmad^{1,2*}, Rabail Urooj^{1,2} and Muhammad Nawaz^{1,2}

*Address all correspondence to: drsaeed@fjwu.edu.pk

1 Department of Environmental Sciences, Fatima Jinnah Women University, Mall Road, Rawalpindi, Pakistan

2 BZ University, Multan, Pakistan

References

- [1] Mulaku C. Mapping and analysis of air pollution in Nairobi, Kenya. International conference on spatial information for sustainable development, Kenya, 2001.
- [2] Gualtieri G and Tartaglia M. Predicting urban traffic air pollution: a GIS framework. *Transportation Research – D*; 1998; 3(5): 329–336.
- [3] Pummakarnchana O, Tripathi N and Dutta J. Air pollution monitoring and GIS modeling: a new use of nanotechnology based solid state gas sensors. *Science and Technology of Advanced Materials* 2005; 6: 251–255.
- [4] Afshar H and Delavar MR. GIS-based air pollution modeling in Tehran. *Environmental Informatics* 2007; 5: 557–566.
- [5] Barnes J, Parsons B and Salter L. GIS Mapping of nitrogen dioxide diffusion tube monitoring in Cornwall, UK. *Air Pollution* 2005; 13: 157–166.
- [6] Elbir T, Mangir N, Kara M, Simsir S, Eren T and Ozdemir S. Development of a GIS-based decision support system for urban air quality management in city of Istanbul. *Atmospheric Environment* 2010; 44: 441–454.
- [7] Veen AVD, Briggs DJ, Collins S, Elliott S, Fischer P, Kingham S, Lebrete E, Pryl K, Reeuwijk HV and Smallbone K. Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science* 2010; 11(7): 699–718.

- [8] Vienneau D, de Hoogh K and Briggs D. A GIS-based method for modeling air pollution exposures across Europe. *Science of the Total Environment* 2009; 408: 255–266.
- [9] Banja M, Como E, Murtaç B and Zotaj A. Mapping air pollution in urban Tirana area using GIS. *International Conference SDI, Skopje*, 15–17 September 2010, 105–114.
- [10] Jensen SS. Mapping human exposure to traffic air pollution using GIS. *Journal of Hazardous Material* 1998; 61(3): 385–392.
- [11] Kim JJ, Smorodinsky S, Lipsett M, Singer BC, Hodgson AT and Ostro B. Traffic-related air pollution near busy roads. *American Journal of Respiratory and Critical Care Medicine* 2004; 170(5): 520–526.
- [12] Alexander SM. Data mining 2005. http://www.eco.utexas.edu/~norman/BUS_FOR/course.mat/Alex/ (Accessed 12th February 2008).
- [13] United Nation. *Conference on the Human Environment*, Sweden, 1972.
- [14] United Nation. *Environmental Performance Annual Report*, New York, 2001.
- [15] US-EPA. Health affects of different air quality index (AQI) levels caused by nitrogen dioxide, 2008.
- [16] Miller D. *Potential hazards of future volcanic eruptions*. California, 1989.
- [17] Atkins DHF, Sandallas J, Law DV, Hough AM and Stevenson K. The measurement of nitrogen dioxide in the outdoor environment using passive diffusion tube samplers. *AEA Technology*, 1986.
- [18] Varshney CK and Singh AP. Passive samplers for NO_x monitoring: a critical review. *The Environmentalist* 2003; 23: 127–136.
- [19] Palmes ED, Gunnison AF, Dimattio J, and Tomczyk C. Personal sampler for nitrogen dioxide. *American Industrial Hygiene Association Journal* 1976; 37: 570–577.
- [20] Gilbert NL, Goldberg MS, Beckerman B, Brook JR and Jerrett M. Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *Journal of the Air & Waste Management Association* 2005; 65: 1059–1063.
- [21] Lozano A, Usero J, Vanderlinden E, Ruez J, Contreras J, Navarrete B and Bakouri HEI. Air quality monitoring network design to control nitrogen dioxide and ozone applied in Granada Spain. *Ozone: Science & Engineering* 2011; 33(1): 80–89.
- [22] Heywood JB. *Internal combustion engine fundamentals*. McGraw-Hill, New York, 1998.

